

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: 08137874 A

(43) Date of publication of application: 31 . 05 . 96

(51) Int. Cl

G06F 17/27
// G06F 9/44

(21) Application number: 06278966

(22) Date of filing: 14 . 11 . 94

(71) Applicant: ATR TSUSHIN SYST
KENKYUSHO:KK

(72) Inventor: NISHIMOTO KAZUSHI

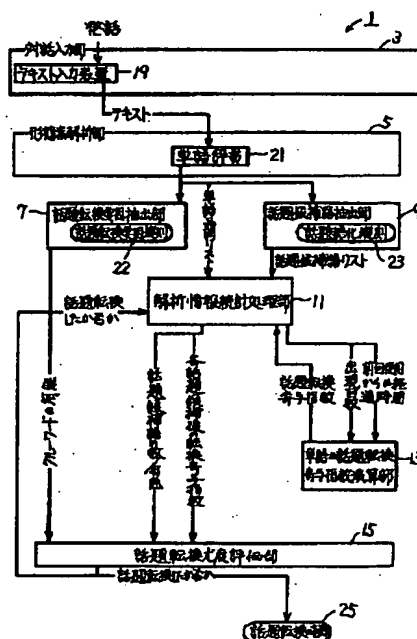
(54) CONVERSATION PROCESSOR

(57) Abstract:

PURPOSE: To provide the conversation processor which can detect whether or not the subject is changed almost in real time without a discrimination between an object oriented type and a free expansion type.

CONSTITUTION: The conversation processor 1 makes a speech of a speaker into a text by the text input device 19 of a conversation input part 3, and converts the text into a list of parts of speech of words by using the word dictionary 21 of a morpheme analysis part 5, a subject candidate word extraction part 9 extracts subject candidate words from the word part-of-speech list, and an analytic information statistical processing part 11 performs statistical processing for the obtained word part-of-speech list and subject candidate word list; and a subject change contribution index arithmetic part 13 for words finds the change contribution index of the subject candidate words, a subject change factor extraction part 7 extracts subject change factors, and a subject change evaluation part 15 evaluates the obtained information to detect whether or not the subject is changed 25.

COPYRIGHT: (C)1996,JPO



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-137874

(43) 公開日 平成8年(1996)5月31日

(51) Int.Cl. ⁹	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/27				
// G 0 6 F 9/44	5 7 0 C	7737-5B 8420-5L	G 0 6 F 15/ 38	M

審査請求 有 請求項の数10 O L (全 19 頁)

(21) 出願番号 特願平6-278966

(22) 出願日 平成6年(1994)11月14日

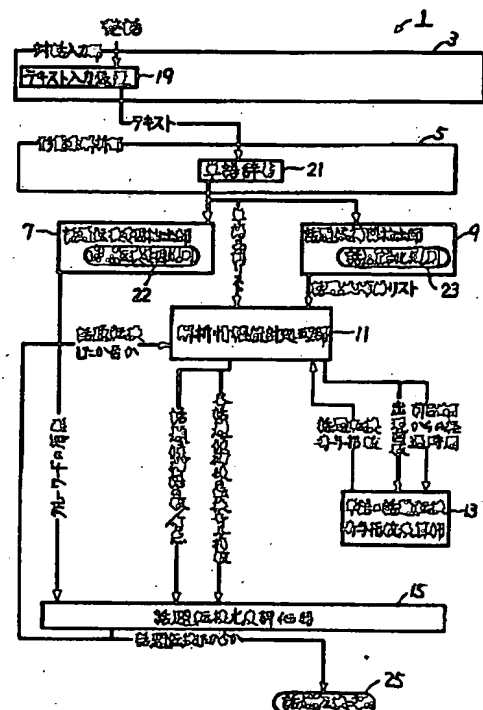
(71) 出願人 000127695
株式会社エィ・ティ・アール通信システム
研究所
京都府相楽郡精華町大字乾谷小字三平谷5
番地
(72) 発明者 西本 一志
京都府相楽郡精華町大字乾谷小字三平谷5
番地 株式会社エィ・ティ・アール通信シ
ステム研究所内
(74) 代理人 弁理士 深見 久郎 (外2名)

(54) 【発明の名称】 対話処理装置

(57) 【要約】

【目的】 目的指向型または自由展開型を問わず、ほぼ実時間で話題転換の有無を検出することができるような対話処理装置を提供する。

【構成】 対話処理装置1は、対話者の発話を対話入力部3のテキスト入力装置19でテキスト化し、形態素解析部5の単語辞書21を用いて単語品詞リストに変換し、その単語品詞リストから話題候補語抽出部9で話題候補語を抽出し、得られた単語品詞リストおよび話題候補語リストに対して解析情報統計処理部11で統計処理を施し、さらに単語の話題転換寄与指数演算部13で話題候補語の転換寄与指数を求め、一方で話題転換要因抽出部7で話題転換要因を抽出し、得られた情報を話題転換評価部15で評価して話題転換有無25を検出する。



BEST AVAILABLE COPY

【特許請求の範囲】

【請求項 1】 複数の話者間での対話内における話題転換の有無を検出する対話処理装置であって、

装置本体と、

それぞれの話者の発話内容を前記装置本体に入力する対話入力手段と、

前記対話入力手段が入力した発話内容をその形態素で解析する形態素解析手段と、

予め定められた話題転換要因規則を前記形態素解析手段の解析結果に適用し、話題転換要因を抽出する話題転換要因抽出手段と、

予め定められた話題化規則を前記形態素解析手段の解析結果に適用し、話題候補語を抽出する話題候補語抽出手段と、

前記形態素解析手段の解析結果および前記話題候補語抽出手段の抽出結果を統計処理するとともに、少なくとも前記話題候補語抽出手段が抽出した話題候補語の転換寄与指数を求める演算処理手段と、

前記対話入力手段が前記装置本体に入力した対話者の発話内容に話題転換が存在するか否かを検出するために、少なくとも前記話題転換要因抽出手段が抽出した話題転換要因、前記演算処理手段の統計処理結果、および前記演算処理手段が求めた話題候補語の転換寄与指数に基づいて話題転換の尤度を評価する評価手段とを備えた、対話処理装置。

【請求項 2】 前記対話入力手段は、それぞれの対話者の発話内容をテキスト形式に変換して前記装置本体に入力するテキスト入力装置を含む、請求項 1 記載の対話処理装置。

【請求項 3】 前記対話入力手段は、いずれの対話者が発話しているかを監視する話者監視装置を含み、前記演算処理手段は、前記話者監視装置の監視結果をも統計処理する、請求項 2 記載の対話処理装置。

【請求項 4】 前記対話入力手段は、それぞれの対話者が発話しているその開始時刻からその終了時刻間の時間を監視する時間監視装置を含み、

前記演算処理手段は、前記時間監視装置の監視結果をも統計処理する、請求項 2 または 3 記載の対話処理装置。

【請求項 5】 前記形態素解析手段は、予め定められた単語辞書を用いて前記対話入力手段が入力した発話内容をその形態素としての単語の品詞で解析する、請求項 1 記載の対話処理装置。

【請求項 6】 前記形態素解析手段は、予め定められた同義語辞書を用いて前記対話入力手段が入力した発話内容をその形態素としての同義語で解析し、その解析結果を前記話題転換要因抽出手段に与える、請求項 5 記載の対話処理装置。

【請求項 7】 前記形態素解析手段は、予め定められた対比語辞書を用いて前記対話入力手段が入力した発話内容をその形態素としての対比語で解析し、その解析結果

を前記演算処理手段が統計処理するために前記演算処理手段に与える、請求項 5 または 6 記載の対話処理装置。

【請求項 8】 さらに、前記評価手段の評価結果に応じて、前記話題候補語抽出手段が抽出した話題候補語に基づく話題語を同定する話題語同定手段を備えた、請求項 1 記載の対話処理装置。

【請求項 9】 前記形態素解析手段は、予め定められた同義語辞書を用いて前記対話入力手段が入力した発話内容をその形態素としての同義語で解析し、

前記話題語同定手段は、前記形態素解析手段の解析結果である同義語にも基づく話題語を同定する、請求項 8 記載の対話処理装置。

【請求項 10】 前記形態素解析手段は、予め定められた対比語辞書を用いて前記対話入力手段が入力した発話内容をその形態素としての対比語で解析し、

前記話題語同定手段は、前記形態素解析手段の解析結果である対比語にも基づく話題語を同定する、請求項 8 または 9 記載の対話処理装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 この発明は、対話処理装置に関し、特に、対話者間での対話内における話題転換の有無を検出できるような対話処理装置に関する。

【0002】

【従来の技術および発明が解決しようとする課題】 従来から行なわれている話題認識技術の大半は、「目的指向型の対話」の話題理解に対するものである。ここで目的指向型の対話とは、たとえば会社の受付での対話、学会への参加申込における対話が挙げられる。このような目的指向型の対話に対して、たとえばプラン・ゴールモデルなどで対話の構造を記述することによって、話題理解が行なわれる。

【0003】ところが、プラン・ゴールモデルが使用される場合には、状況の数だけスクリプトが準備されなければならない。したがって、目的指向型の対話の話題理解を行なえる装置の柔軟度が欠けることとなる。さらに、このような装置では、特に目的を持たない「自由展開型の対話」には適用されず、自由展開型の対話の話題理解が行なわれない。

【0004】一方、わずかに自由展開型対話に対する話題理解の装置が存在する。図 9 は、そのようなわずかに自由展開型対話に対する話題理解の装置での話題理解の方法を説明するための図である。

【0005】図 9 を参照して、まず、対話者間で対話が行なわれているとする。それによって、発話 311、発話 313、発話 315、発話 317、発話 319 のように、この順で発話が得られたとする。次に、発話 311 をコンピュータ処理可能なテキスト 301 に変更し、発話 313 をテキスト 303 に変更し、発話 315 をテキスト 305 に変更し、発話 317 をテキスト 307 に変

更し、発話319をテキスト309に変更する。テキスト301には話題候補語321a, 321bが含まれている。テキスト303には、話題候補語323a, 323bが含まれる。テキスト305には、話題候補語325が含まれる。テキスト307には、話題候補語327a, 327bが含まれる。テキスト309には、話題候補語329が含まれる。

【0006】このようにして得られた5発話311, 313, 315, 317, 319に対応するテキスト301, 303, 305, 307, 309が、話題スタック331にスタックされる。スタックされたテキスト301, 303, 305, 307, 309には、話題候補語321a, 321b, 323a, 323b, 325, 327a, 327b, 329が含まれている。そこで、これらの話題候補語321a, 321b, 323a, 323b, 325, 327a, 327b, 329の中で、同一の話題候補語が存在する場合には、話題が同定される。さらに、その話題同定後に、話題転換が検出される。このようにして、自由展開型に対する話題理解が行なわれる。

【0007】しかしながら、このような自由展開型対話に対する話題理解の方法が存在するとはいえ、実際には、対話の内容に応じた分野辞書が準備される必要がある。したがって、準備された分野から外れた内容の話題理解は行なわれず、本当の意味での自由展開型対話に対する話題理解は行なわれていない。

【0008】さらに、話題同定のために話題スタックが使用されているため、話題転換の認識、話題同定に数発話の遅れが発生する。前述の例では、5発話の遅れが発生する。したがって、リアル時間に近い状態での話題の転換の認識、話題同定は行なわれない。

【0009】以上のことをまとめる。従来話題理解では、目的指向型の対話であれ、自由展開型の対話であれ、分野ごとに準備されたドメイン知識から外れた分野の話題には適応されない。さらに、従来の自由展開型の対話では、まず話題同定を行ない、その後話題転換を認識するため、実時間での話題転換の検出は行なわれない。

【0010】ゆえに、本発明の目的は、ほぼ実時間で話題転換を検出でき、目的指向型の対話および自由展開型の対話におけるいずれの分野の話題にも変更されことなく適用可能な対話処理装置を提供することである。

【0011】

【課題を解決するための手段】請求項1の発明に係る対話処理装置は、複数の話者間での対話内における話題転換の有無を検出する対話処理装置であって、装置本体と、それぞれの話者の発話内容を装置本体に入力する対話入力手段と、対話入力手段が入力した発話内容をその形態素で解析する形態素解析手段と、予め定められた話題転換要因規則を形態素解析手段の解析結果に適用し、

話題転換要因を抽出する話題転換要因抽出手段と、予め定められた話題化規則を形態素解析手段の解析結果に適用し、話題候補語を抽出する話題候補語抽出手段と、形態素解析手段の解析結果および話題候補語抽出手段の抽出結果を統計処理するとともに、少なくとも話題候補語抽出手段が抽出した話題候補語の転換寄与指数を求める演算処理手段と、対話入力手段が装置本体に入力した対話者の発話内容に話題転換が存在するかどうかを検出するために、少なくとも話題転換要因抽出手段が抽出した話題転換要因、演算処理手段の統計処理結果、および演算処理手段が求めた話題候補語の転換寄与指数に基づいて話題転換の尤度を評価する評価手段とを備えて構成される。

【0012】請求項2の発明に係る対話処理装置では、請求項1の対話入力手段は、それぞれの対話者の発話内容をテキスト形式に変換して装置本体に入力するテキスト入力装置を含む。

【0013】請求項3の発明に係る対話処理装置では、請求項2の対話入力手段は、いずれの対話者が発話しているかを監視する話者監視装置を含み、演算処理手段は、話者監視装置の監視結果をも統計処理する。

【0014】請求項4の発明に係る対話処理装置では、請求項2または3の対話入力手段は、それぞれの対話者が発話しているその開始時刻からその終了時刻間の時間を監視する時間監視装置を含み、演算処理手段は、時間監視装置の監視結果をも統計処理する。

【0015】請求項5の発明に係る対話処理装置では、請求項1の形態素解析手段は、予め定められた単語辞書を用いて対話入力手段が入力した発話内容をその形態素としての単語の品詞で解析する。

【0016】請求項6の発明に係る対話処理装置では、請求項5の形態素解析手段は、予め定められた同義語辞書を用いて対話入力手段が入力した発話内容をその形態素としての同義語で解析し、その解析結果を話題転換要因抽出手段に与える。

【0017】請求項7の発明に係る対話処理装置では、請求項5または6の形態素解析手段は、予め定められた対比語辞書を用いて対話入力手段が入力した発話内容をその形態素としての対比語で解析し、その解析結果を演算処理手段が統計処理するために演算処理手段に与える。

【0018】請求項8の発明に係る対話処理装置は、さらに、評価手段の評価結果に応じて、話題候補語抽出手段が抽出した話題候補語に基づく話題語を同定する話題語同定手段を備えて構成される。

【0019】請求項9の発明に係る対話処理装置では、請求項8の形態素解析手段は、予め定められた同義語辞書を用いて対話入力手段が入力した発話内容をその形態素としての同義語で解析し、話題語同定手段は、形態素解析手段の解析結果である同義語にも基づく話題語を同

定する。

【0020】請求項10の発明に係る対話処理装置では、請求項8または9の形態素解析手段は、予め定められた対比語辞書を用いて対話入力手段が入力した発話内容の形態素としての対比語で解析し、話題語同定手段は、形態素解析手段の解析結果である対比語にも基づく話題語を同定する。

【0021】

【作用】請求項1の発明に係る対話処理装置は、対話者間の対話の内容に依存する構成を有していないため、い

10

ずれの分野の話題にも適用でき、さらに、話題同定に数発話を蓄積する必要がないためほぼ実時間に近い状態で話題転換の有無を検出できる。

【0022】請求項2の発明に係る対話処理装置は、対話入力手段としてテキスト入力装置を用いることで、コンピュータ処理可能なテキスト形式に変換できる。

【0023】請求項3の発明に係る対話処理装置は、対話入力手段として話者監視装置を含むので、話者指定による話題転換の発生の検出が可能となり、話題転換検出精度を向上させることができる。

20

【0024】請求項4の発明に係る対話処理装置は、対話入力手段として時間監視装置を含むので、発話間の空白期間による話題転換の発生を検出でき、話題転換検出精度を向上させることができる。

【0025】請求項5の発明に係る対話処理装置は、形態素解析手段が単語の品詞によって発話内容を解析するので、単語の品詞別の話題候補を抽出できる。

【0026】請求項6の発明に係る対話処理装置は、発話内容を同義語でも解析するので、同義語から得られる話題転換要因を抽出できる。

30

【0027】請求項7の発明に係る対話処理装置は、発話内容を対比語でも解析するので、その対比語から得られる解析結果を統計処理することで、対比語に関する統計処理情報を得ることができる。

【0028】請求項8の発明に係る対話処理装置は、話題転換が存在するか否かを検出するための評価結果に応じて話題候補語に基づく話題語を同定するので、請求項1の作用に加えて、ほぼ実時間に近い話題同定を行なえる。

【0029】請求項9の発明に係る対話処理装置は、発話内容を同義語で解析し、同義語にも基づく話題語を同定するので、話題同定の精度を向上させることができる。

40

【0030】請求項10の発明に係る対話処理装置は、発話内容を対比語で解析し、対比語にも基づく話題語を同定するので、話題同定の精度を向上させることができる。

【0031】

【実施例】図1は、この発明の一実施例による対話処理装置の概略ブロック図である。

【0032】図1を参照して、人々の間で行なわれる対話を処理する対話処理装置1は、対話入力部3と、形態素解析部5と、話題転換要因抽出部7と、話題候補語抽出部9と、解析情報統計処理部11と、単語の話題転換寄与指数演算部13と、話題転換尤度評価部15を含む。

【0033】対話入力部3は、対話を行なっている人々のそれぞれの発話を入力する。対話入力部3は、テキスト入力装置19で構成される。テキスト入力装置19は、人々の発話内容をテキスト形式で装置本体に入力する。たとえば、テキスト入力装置19が音声認識装置で構成された場合には、人々のうちの話者の音声による発話に対して音声認識を行ない、その発話内容をテキスト形式に変換して形態素解析部5に出力する。また、テキスト入力装置19には、キーボードなどが用いられてもよく、用いられたキーボードによって話者の発話内容がテキスト形式に変換されて形態素解析部5に与えられてもよい。さらに、テキスト入力装置19は、人々の間で行なわれる手話や点字に対しての対話の発話内容を入力してもよい。

【0034】形態素解析部5は、単語辞書21を有している。そして、単語辞書21を用いて、形態素解析部5は、対話入力部3でテキスト化された発話文を単語に分解する。その後、形態素解析部5は、分解された単語の品詞を決定する。そして、形態素解析部5は、単語品詞リストを話題転換要因抽出部7、話題候補語抽出部9および解析情報統計処理部11に与える。

【0035】話題転換要因抽出部7は、予め定められた話題転換要因規則22を知識として有している。そして、形態素解析部5が決定した単語の品詞に転換要因規則22を適用して、話題転換を発生させやすい単語の検出を行なう。たとえば、そのような単語としては、キーワード（話変わって、ところで、など）、疑問文、または相手を特定する語もしくは言回しなどが挙げられる。そして、話題転換要因抽出部7は、たとえばキーワードの有無を話題転換尤度評価部15に与える。

【0036】話題候補語抽出部9は話題化規則23を知識として有している。そして、形態素解析部5が決定した単語の品詞にその話題化規則23が適用されて、規則に当てはまる構文が発話中から検出される。そして、検出が行なわれたことで、その構文を指示する話題候補が得られ、話題候補語リストが解析情報統計処理部11に出力される。

【0037】解析情報統計処理部11は、形態素解析部5、話題候補語抽出部9、後で説明する単語の話題転換寄与数演算部13、および後で説明する話題転換尤度評価部15で得られる情報を基に、統計処理を行なう。そして、統計処理後に得られる各話題候補語の転換寄与指数および話題候補語の数／有無を話題転換尤度評価部15に与える。

50

【0038】単語の話題転換寄与指数演算部13は、解析情報統計処理部11で得られる、入力された名詞の出現回数、最も最近にその語が使用されてからの経過時間に基づいて、所定の方法によってその語の話題転換寄与指数を演算して解析情報統計処理部11に出力する。

【0039】話題転換尤度評価部15は、解析情報統計処理部11が統計処理することで得られた話題候補語の数／有無および各話題候補語の転換寄与指数と、話題転換要因抽出部7で得られたたとえばクルーワードの有無に基づいて、現在の発話において話題転換が発生しているかどうかを評価する。そして、その評価結果によって話題転換の有無25が得られる。また、話題転換したか否かは、解析情報統計処理部11にも与えられる。

【0040】次に、動作について簡単に説明する。まず、対話入力部3に入力された発話が、テキスト入力装置19によってコンピュータ処理可能な形式(テキスト)に変換される。そして、変換されたテキストは形態素解析部5に与えられる。

【0041】次に、形態素解析部5は、単語辞書21を使用して、入力された発話を解析して単語に分解し、各単語の品詞を決定する。解析結果である単語品詞リストは、話題転換要因抽出部7、話題候補語抽出部9および解析情報統計処理部11に与えられる。

【0042】次に、話題転換要因抽出部7は、形態素解析部5の解析結果中に話題の転換を促す語(句)が現れているかどうかを話題転換要因規則22に従ってチェックする。そのような語(句)が検出されれば、話題転換要因抽出部7は解析情報統計処理部11に通知する。

【0043】話題候補語抽出部9は、形態素解析部5の解析結果に話題化規則23を適用し、規則に一致する構文が発見されれば、その構文が指示する話題候補語を抽出して、解析情報統計処理部11に与える。

【0044】解析情報統計処理部11は、対話入力部3、形態素解析部5、話題転換要因抽出部7、話題候補語抽出部9から入力される情報に基づき、単語／文／発話単位に各量の統計を取る。また、形態素解析部5および話題候補語抽出部9のそれぞれから名詞および話題候補語が入力されて、その後の出現回数と、初出語でない場合は前回に出現してから経過時間を単語の話題転換寄与指数演算部13に与える。そして、単語の話題転換寄与指数演算部13の演算結果を受取って、その値も統計処理する。

【0045】単語の話題転換寄与指数演算部13は、解析情報統計処理部11から入力されるある語の出現回数と前回に出現してから経過時間を用いて、その語が話題の転換発生に寄与する度合を計算し、結果を解析情報統計処理部11に与える。

【0046】話題転換尤度評価部15は、解析情報統計処理部11からたとえば話題転換候補語の数／有無や各話題候補語の転換寄与指数のような各種の統計情報、話

題転換要因抽出部7から与えられるたとえばクルーワードの有無のような話題転換要因に基づき、所定の評価手法によって現在の発話において話題転換があったかどうかを評価する。その評価結果は、解析情報統計処理部11に与えられる。また、話題転換尤度評価部15は、話題の転換ありと判断した場合には、その旨をたとえば利用者に通知する。

【0047】このような装置によって、目的指向型の対話のみならず、特に目的を持たない自由展開型の対話に適用可能である話題転換の有無を検出する対話処理装置が提供される。また、モデルや話題分野に応じた辞書などの対話の内容に依存する部分が処理の中で存在しないため、どんな分野の話題に対しても変更する必要のない装置が提供される。これは、対話の構造、内容の分野などが一切仮定されず、話題の転換の検出が実際の人間から獲得されたヒューリスティックなルールに基づいて行なわれているためである。さらに、話題転換および話題の検出が時間遅れなく行なわれる。すなわち、従来例のように話題転換の認識・話題同定に話題スタックなどが使用されておらず、現在の発話とそこまでの発話との差が検出されることで変化が把握されているため、ほぼ実時間に近い話題転換の検出が行なわれる。

【0048】図2は、この発明の他の実施例による対話処理装置を示した概略ブロック図である。

【0049】図2を参照して、この実施例の対話処理装置101は、対話入力部103と、形態素解析部105と、話題転換要因抽出部107と、話題候補語抽出部109と、解析情報統計処理部111と、単語の話題転換寄与指数演算部113と、話題転換尤度評価部115と、話題語同定部117を含む。

【0050】図1に示した実施例と異なる部分について特に説明する。対話入力部103は、テキスト入力装置19に加え、話者監視装置119と、発話推移時間監視装置121を含む。話者監視装置119には、人々の発話が入力されており、話者監視装置119は現在の話者が誰であるかを監視して、その監視結果を示す話者IDを解析情報統計処理部111に与える。発話推移時間監視装置121には、人々の発話が入力されており、発話推移時間監視装置121は発話行為に係る時間情報を獲得する。そのような時間情報としては、現在の発話の開始時刻と、現在の発話の終了時刻とが挙げられる。これらの情報は解析情報統計処理部111に与えられる。

【0051】次に、形態素解析部105は、単語辞書21に加え、対比語辞書123と、同義語辞書125とを有する。したがって、形態素解析部105は、対比語辞書123を用いて対比語(反意語)を検出する。同様に、形態素解析部105は、同義語辞書125を用いて同義語を検出する。そして、形態素解析部105の対比語辞書123で得られた対比語リストは、解析情報統計処理部111および後で説明する話題同定部117に与

えられる。さらに、形態素解析部 105 の同義語辞書 125 で得られた同義語リストは、話題転換要因抽出部 107 および話題語同定部 117 に与えられる。

【0052】話題転換要因抽出部 107 は、話題転換要因規則 22 を用いて、以下のような話題転換要因を抽出して話題転換尤度評価部 115 に与える。すなわち、図 1 の実施例で示したキーワードの有無のみならず、疑問文か否か、相手特定語の有無、同義語の数/有無、指示語の数/有無、接続詞の数/有無を話題転換尤度評価部 115 に与える。キーワードの有無、疑問文か否か、相手特定語の有無、指示語の数/有無および接続詞の数/有無は、単語辞書 21 によって与えられた単語品詞リストから、得られる。同義語の数/有無は、同義語辞書 125 から得られた同義語リストによって得られる。

【0053】話題候補語抽出部 109 は、図 1 の話題候補語抽出部 9 とほぼ同様であり、得られた話題候補語リストを解析情報統計処理部 111 だけでなく、話題語同定部 117 にも与える。

【0054】次に、解析情報統計処理部 111 は、図 1 の解析情報統計処理部 11 よりも多くの統計処理結果を得て、それを話題転換尤度評価部 115 に与える。まず、与えられた単語品詞リストに従って、解析情報統計処理部 111 は、話題候補語の数/有無、各話題候補語の転換寄与指数だけでなく、初出語の数/有無、初出語かつ話題候補語の数/有無、各初出語の転換寄与指数を話題転換尤度評価部 115 に与える。また、対話入力部 103 に話者監視装置 119 および発話推移時間監視装置 121 が加わったこと、形態素解析部 105 に対比語辞書 123 が加わったことに伴って、解析情報統計処理部 111 は以下のような統計情報を得て話題転換尤度評価部 115 に与える。

【0055】まず、形態素解析部 105 に対比語辞書 123 が加わったことに伴って、前文対比語の数/有無の情報が得られ、それを話題転換尤度評価部 115 に与える。対話入力部 103 に話者監視装置 119 および発話推移時間監視装置 121 が加わったことに伴って、解析情報統計処理部 111 は、話者の交代の有無、前話題転換からの経過時間、現発話者の前発話からの経過時間、現発話の発話時間、および前発話終了から現発話開始までの時間に関する情報を得て、それを話題転換尤度評価部 115 に与える。

【0056】次に、単語の話題転換寄与指数演算部 113 は、解析情報統計処理部 111 が話題転換尤度評価部 115 に各話題候補の転換寄与指数だけでなく、各初出語の転換寄与指数をも与えるため、この各初出語の転換寄与指数を演算して求め解析情報統計処理部 111 に与える。

【0057】次に、話題転換尤度評価部 115 は、話題転換要因抽出部 107 から与えられた話題転換要因、解

析情報統計処理部 111 から与えられた統計処理結果の情報を含めて評価し、話題転換したか否かを決定する。

【0058】次に、話題語同定部 117 は、話題転換尤度評価部 115 が話題転換があると評価したことに応じて、形態素解析部 105 の対比語辞書 123 および同義語辞書 125 によって得られた対比語リストおよび同義語リストと、話題候補語抽出部 109 によって得られた話題候補語リストとに従って、話題語を同定し、話題 127 を利用者に提供する。

10 【0059】このような新たな構成を加えた対話処理装置 101 では、図 1 に示す実施例と同様に、話題転換のほぼ実時間による検出が可能となる。さらに、新たな構成によって以下のような効果が得られる。

【0060】まず、話者監視装置 119 によって、話者指定による話題転換の発生のルールが適用可能となり、転換検出精度が向上する。次に、発話推移時間監視装置 121 によって、発話と発話の間の空きかたによる話題転換の発生のルールが適用可能となり、転換検出精度が向上する。次に、同義語辞書 125 および対比語辞書 123 の使用により、同義語および対比語による話題転換のルールが適用可能となり、転換検出精度が向上する。次に、話題語同定部 117 の追加によって、話題転換検出時に転換した話題を表わす語の抽出が可能となる。

【0061】なお、話題語同定部 117 が同定した結果が解析情報統計処理部 111 に与えられてもよい。そして与えられた話題語が何度話題化に寄与したかを示す話題化回数が解析情報統計処理部 111 で統計処理されれば、より話題転換検出精度が向上する。

30 【0062】次に、このような効果をもたらす本装置の動作について詳しく説明する。まず、対話入力部 103 のテキスト入力装置 19 は、入力された発話をコンピュータ処理可能な形式(テキスト)に変換し、形態素解析部 105 の対比語辞書 123、単語辞書 21 および同義語辞書 125 に与える。また、話者監視装置 119 によって現在の話者が監視される。そして、発話推移時間監視装置 121 は、発話行為にかかる推移時間情報を検出する。そして、話者監視装置 119 の出力である検出結果(話者 id) および話者推移時間監視装置の検出結果が解析情報統計処理部 111 に与えられる。

40 【0063】形態素解析装置 105 は、同義語辞書 125 および対比語辞書 123 を使用して、同義語および対比語の検出を行なう。その検出された結果である同義語リストは話題転換要因抽出部 107 および話題語同定部 117 に与えられ、対比語リストは解析情報統計処理部 111 および話題語同定部 117 に与えられる。

【0064】話題転換要因抽出部 107 は、図 1 の実施例で示したように形態素解析部 105 の単語辞書 21 から与えられた単語品詞リストに従って、話題転換要因として以下の情報を検出する。その以下の情報とは、第 1 はキーワードの有無であり、第 2 は疑問文か否かであ

り、第3は相手特定語の有無であり、第4は同義語（同語）の数／有無であり、第5は、指示語の数／有無であり、第6は、接続詞の数／有無である。

【0065】第1のキーワードとは、一般に明示的に話題の転換を促す語（句）であり、その例としては、ところで、次に、話変わって、それから、後、等が挙げられる。このキーワードの有無は、話題転換要因抽出部107に与えられる単語品詞リストから、話題転換要因規則22に記述されるキーワードに相当する語が見つけ出されることで得られる。

【0066】第2の疑問文か否かは、疑問文が出現した場合に話題転換として認識されやすいという、実験に基づくヒューリスティックなルールによって得られる情報である。この疑問文か否かは、話題転換要因抽出部107に与えられる単語品詞リストから、話題転換要因規則22に記述される疑問文に相当するかどうか判断されることで得られる。

【0067】第3の相手特定語の有無は、対話に参加している特定の人に向かって話しかける場合に話題転換として認識されやすいという、実験に基づくヒューリスティックなルールによって得られる情報である。たとえば、参加者の名前、人称代名詞が挙げられる。この相手特定語の有無は、話題転換要因抽出部107に与えられる単語品詞リストから、話題転換要因規則22に記述される相手特定語に相当する語が取出されることで得られる。

【0068】第4の同義語の数／有無は、1発話または1文中に同義語（同語）が出現する場合、それらの語は強調されている（つまり話題を示す語である）と認識されやすいという、実験に基づくヒューリスティックなルールによって得られる情報である。この同義語の数／有無は、話題転換要因抽出部107に与えられる同義語リストから同義語が取出されることで得られる。

【0069】第5の指示語の数／有無は、前発話に関する指示語が出現する発話は話題が継続しているものと認識しやすいという実験に基づくヒューリスティックなルールによって得られる情報である。この指示語の数／有無は、話題転換要因抽出部107に与えられる単語品詞リストから指示語が取出されることで得られる。

【0070】第6の接続詞の数／有無は、前発話に対する接続詞が出現する発話は話題が継続しているものと認識されやすいという実験に基づくヒューリスティックなルールによって得られる情報である。この接続詞の数／有無は、話題転換要因抽出部107に与えられる単語品詞リストから接続詞が取出されることで得られる。

【0071】次に、話題候補語抽出部109は、話題化規則23によって話題候補語を抽出する。その話題規則23としては、以下のようなものが挙げられる。

【0072】まず第1は、強調文の名詞が話題候補語とされる。強調文の例としては名詞＋助動詞「だ」、また

は名詞＋助動詞「で」が挙げられる。第2は、体言止めの名詞が話題候補語とされる。たとえば、名詞＋句点「。」の構文の場合の名詞が話題候補語とされる。

【0073】第3は、名詞＋格助詞＋提題助詞の構文の場合の名詞が話題候補語とされる。第4は、名詞＋提題助詞の構文の場合の名詞が話題候補語とされる。第5は、名詞＋副助詞＋提題助詞の構文の場合の名詞が話題候補語とされる。第6は、名詞＋取立助詞の構文の場合の名詞が話題候補語とされる。第7は、名詞＋格助詞＋取立助詞の構文の場合の名詞が話題候補語とされる。

【0074】第8は、存在文の場合の名詞が話題候補語とされる。存在文の例としては、名詞＋格助詞「が」＋ある（本動詞語幹）、名詞＋格助詞「が」＋い（本動詞語幹）、または名詞＋格助詞「が」＋あ（本動詞語幹）の構文が挙げられる。

【0075】第9は、名詞＋副助詞＋取立助詞の構文の場合の名詞が話題候補語とされる。第10は、名詞＋副助詞「とか」の構文の場合の名詞が話題候補語とされる。

【0076】このような話題候補語をも入力される解析情報統計処理部111は、以下のように管理表を作成して情報を管理する。

【0077】図3は、そのような管理を説明するための第1の図であり、図4は、そのような管理を説明するための第2の図であり、図5は、そのような管理を説明するための第3の図である。

【0078】図3から図5を参照して、解析情報統計処理部111は、図3の対話管理表201で管理する。対話管理表201には、カレント・メタ情報203と、話者情報205と、言語情報207と、文脈情報209とが含まれる。カレント・メタ203には、発話番号、文番号、発話内文番号および発話IDが含まれる。

【0079】話者情報205は、話者IDを添字とする配列で示されている。たとえば、話者IDが0の情報には、最近発話文番号211と、最近発話時刻213と、全発話回数215と、名前217とが情報として含まれる。このような話者情報に関しては後で具体的に説明する。

【0080】言語情報207には名詞要素リスト219が含まれる。名詞要素リスト219は、図4に示すように、見出し語221、出現回数223、ENTRY（対比語）225と、話題化回数227と、話題認識回数229と、最新使用文番号231と、Left（左）233と、Right（右）235とが含まれる。見出し語221は、シソーラス同義語情報により正規化されている。ENTRY（対比語）225は、反対語を参照するために取入れられている。そして、名詞要素リスト219のLeft（左）233には、さらに名詞要素リスト237が記録され、名詞要素リスト219のRight（右）235には名詞要素リスト239が記録されている。

る。このように、左右の名詞要素リストが各名詞要素リストに記録されることにより、ツリー構造が作られる。これにより、検索が容易となる。このような、名詞要素リストについては後で具体的に説明する。

【0081】次に、図3の対話管理表201に戻って、対話管理表201の文脈情報209には、カレント情報241と、短期発話メモリ(FIFO)243とが含まれる。カレント情報241には、最近転換文番号245と、現在の話題247とが含まれる。

【0082】このようなカレント情報241以外のカレント情報は、短期発話メモリ(FIFO)243に含まれており、短期発話メモリ(FIFO)243は、5つの発話情報によって形成されている。その1つの発話情報249は、図5に示すように、発話番号251と、発話者番号253と、文情報リスト255と、前発話情報ポインタ257とを含む。そして、文情報255は、文番号259と、発話内文番号261と、各種指標(ビット表現)263と、有題文フラグ265と、チャート(形態素情報ほか)267と、NEXT文情報269とを含む。さらに、各種指標(ビット表現)263は、位置と区分と内容によって項分けされた情報271を含む。このような、発話情報249および文情報255については後で説明する。

【0083】次に、表1から表4を用いて具体的に説明する。

【0084】

名詞要素リスト

見出し語	競馬	ジャンル	夫婦単位	男	女
出現回数	3	3	1 → 2	3	2	1	2	1
ENTRY(対比語)	-	-	- → -	-	-	-	-	-
話題化回数	1	2	0 → ?	1	2	0	1	1
話題認識回数	1	0	0 → ?	0	0	0	0	0
最新使用文番号	8	18	2 → 23	21	20	9	16	11
単語評価値(%)	1.30	1.80	1.00 → 1.02	1.20	1.20	1.00	1.50	1.00

【0087】

* 【表1】

対話管理表

カレント・メタ情報	
発話番号	13
文番号	23
発話内文番号	1
話者ID	0
言語情報(名詞要素リスト)	
文脈情報	

【0085】

【表2】

話者情報

話者ID	0	1	2
最近文番号	21	12	22
最近終了時刻	489	245	498
全発話回数	5	3	4
名前	黒鉄	源一郎	直子

【0086】

【表3】

【表4】

発 話 情 報		文 情 報														そ の 他		
		各 種 指 標 (ビット表現)																
		文 節 要 素				文 節 要 素				文 節 要 素				判定	形態素			名 詞 要 素
		U	開 始 時 刻	終 了 時 刻	文	7/4	S	SI	Q	接 指	同 対	初 クル	初 提					
9	↓	14時30分05秒 14時30分10秒	↓	徳馬	13	1		1			1		1			省略	男, 女	
10	↓	14時30分11秒	↓	徳馬	14	1										"	なし	
↓		14時30分20秒	↓	徳馬	15	2					1					"	投資	
11	↓	14時30分22秒	↓	徳馬	16	1					1			1		"	社会学, 遊び, 四本柱, 徳馬, 偶然, キャンブル	
↓			↓	徳馬	17	2				1	1			1		"	四本柱, 順列, 組合わせ, 遊び, 正体	
↓			↓	徳馬	18	3					1			1		"	ギャンブル, 定額, 偶然, 対比, 必然	
↓			↓	徳馬	19	4			1							"	なし	
↓			↓	徳馬	20	5								1		"	女, 必然, 目	
↓		14時30分48秒	↓	徳馬	21	6					1	1		1		"	男, 投資, 偶然性, 不足, 遊び	
12	↓	14時30分44秒 14時30分46秒	↓	徳馬	22	1				1				1		"	偶然と必然	
13	↓	14時30分51秒	↓	お二人 お二人	23	1		1						1	1	"	お二人, 夫婦単位	

【0088】表1に示すような対話管理表が得られているとする。すなわち、発話番号は13であり、文番号は23であり、発話内文番号は1であり、話者IDは0の情報得られているとする。この情報は、以下のような状態を示している。すなわち、話者ID、0（名前は黒鉄）の人が喋っている状態は、ある基準から数えて13番目の発話であり、同じ基準から数えて23番目の文であること、さらに、13番目の発話内における1番目の文であることを示している。

【0089】さらに、表2の話者情報によって、現在発

話を行なっている話者IDが0の人の前に発話を行なった人は、話者IDが2の人であることが最近の文番号の示す値からわかる。また、話者IDが0、1および2の全員の全発話回数を加えれば、12であることより、表1の対話管理表における発話番号が13であることも明らかである。そして、表2の話者情報の各話者IDごとの情報は、図3の話者情報205と対応している。

【0090】次に、表1の対話管理表の言語情報（名詞要素リスト）は、具体的には表3に示されるような情報である。表3の各見出し語に対応した情報が、図4の名

詞要素リスト219, 237, 239のそれぞれに対応している。特に、話題化回数は、話題候補語抽出部109によって得られ、話題認識回数は、話題語同定部117によって得られる。

【0091】ところで、表3には単語評価値(%)が記載されているが、図4の名詞要素リストにはそのような記載はない。これは、単語の話題転換寄与指数演算部113で得られる話題転換寄与指数が解析情報統計処理部111にフィードバックされてもまたはされなくてもよいことを示すためである。すなわち、表3で得られる名詞要素リストの場合には、単語の話題転換寄与指数演算部113からの話題転換寄与指数のフィードバックがあった場合であり、図4に示す名詞要素リストの場合には単語の話題転換寄与指数演算部113からの話題転換寄与指数がフィードバックされず、話題転換尤度評価部115に直接与えられた場合である。したがって、単語の話題転換寄与指数演算部113の出力が解析情報統計処理部111にフィードバックされない場合には、各話題候補語の転換寄与指数および各初出語の転換寄与指数は単語の話題転換寄与指数演算部113から話題転換尤度評価部115に直接入力される。話題転換寄与指数のフィードバックが必要な場合としては、話題転換寄与指数を求める関数が漸化式で表現されており、以前の話題転換寄与指数の値が現在の話題転換寄与指数の計算のために必要な場合などが挙げられる。

【0092】次に、対話管理表の文情報は、具体的には表4に示すような情報である。表4に示す文脈情報は、発話情報と文脈情報に大きく分けて別れている。

【0093】発話情報には、Uで表わされる発話番号と、各発話の開始時刻および終了時刻と、転で表わされる話題が転換した文番号と、ワダイとが示されている。発話番号U=9は、開始時刻が14時30分05秒であり、終了時刻が14時30分10秒である。発話番号U=10の開始時刻は14時30分11秒であり、終了時刻は14時30分20秒である。発話番号U=11の開始時刻は14時30分22秒であり、終了時刻は14時30分43秒である。発話番号U=12の開始時刻は14時30分44秒であり、終了時刻は14時30分46秒である。そして、発話番号U=13の開始時刻は14時30分51秒である。発話情報の最新の入力文は、発話番号が13、話題転換の文番号が23、話題がお二人となっており、表1の対話管理表に対応している。

【0094】文脈情報には、メタ情報と、各種指標(ビット表現)と、その他の情報が示される。メタ情報には、表でSで表わされる文番号と、SUで表わされる発話内の文番号が示されている。最新の入力文である文番号が23、発話内の文番号が1となっており、表1の対話管理表と対応している。

【0095】各種指標(ビット表現)には、後続要因、前節要因、転換要因、判定が含まれる。後続要因には、

表でQで表わされる疑問文であるか否かを示す情報が示されている。前節要因には、表でそれぞれ接、指、同、対で表わされるように、接続詞、指示語、同義語、対比語を示す情報が示されている。転換要因には、それぞれ初、クル、初提、相、題で示される、初出語の有無、クルワードの有無、初出語でありかつ話題候補語である語の有無、対話の相手を特定する語の有無、話題候補語の有無を示す情報が示されている。

【0096】判定には、表で転で示す話題の転換があったか否かを示す情報が示されている。発話推移時間監視装置121がある場合に、ここに話題転換が発生した発話の開始時刻が記録される。

【0097】このように表4で示される文脈情報のたとえばメタ情報の文番号が16から21にかけての情報が、図5の発話情報249に対応し、文脈情報の各1行が図5の文脈情報255に対応する。

【0098】なお、図3から図5および表1から表4を用いた解析情報統計処理部では、統計処理する必要のない情報も管理しているが、図2に示すように解析情報統計処理部111から出力する情報に関してのみ管理してもよい。

【0099】以上のように、解析情報統計処理部111は、種々の情報を統計処理する。次に、このような統計処理の結果得られる入力された単語の出現頻度および最近出現してからの経過時間に従って、単語の話題転換寄与指数演算部113は話題転換寄与指数を求める。入力された単語の出現頻度をfで表わし、最近出現してからの経過時間をtで表わすと、第(1)式に示すような関数によって寄与指数hが求められる。また、第(1)式の右辺のg(t)は、第(2)式によって与えられる。

$$【0100】 h = e^{g(t)f} \dots (1)$$

$$g(t) = (1 + kt)^{-1} \dots (2)$$

ここで、kは時間の重みを調整する係数である。このような関数は図6に示されるような関数となる。

【0101】図6におけるグラフでは、平面座標の一方の軸がt(経過時間)であり、他方の軸がf(出現頻度)とされており、さらに縦軸にはh(評価値)が取られている。図6から明らかなように、前に出現してから今回出現するまでの経過時間が長いほど、評価値の値は大きくなっている。これに対して、出現回数が増えるほど、評価値の値は小さくなっている。

【0102】なお、第(1)式で初出語の場合のfの値は0とする。すなわち、初出語は必ず評価値(h)の値が1とされるように設定されている。

【0103】このように、ある単語が初めて出現した場合に、または久し振りに出現した場合に寄与率が高く、逆にしばしば出現する場合、またはほんのちょっと前にも出現したことがある場合に寄与率が低くなるような関数が用いられることで、話題転換の評価が可能となっ

50 いる。

【0104】図7は、第(1)式に示されるような関数
が実際にどのような値を示すかを実験した結果を示した
グラフである。図7において、横軸は出現時刻(t)を
示し、縦軸は出現単語話題転換指数(h)を示す。ただ
し、出現時刻(t)として、実際の経過時間でなく文字
数を使用し、 $k=0.01$ と設定されている。さらに、
ある対談記事中の「男」という単語の寄与指数について
図7に示すグラフは表わしている。

【0105】図7を参照して、 $t=2000$ までは連続
的に「男」という単語が頻出するので、寄与指数は時間
とともに低下している。しかし、 $t=2000$ から50
00の間では発話中に出現していないため、 $t=500$
0付近で出現した際の寄与指数は再び高く評価されてい
る。つまり、 $t=1500$ 辺りでは「男」という単語に
よる話題転換はあまり促進されていないが、 $t=500$
0付近で再出現した際には「男」によって話題が転換す
る可能性が高いことを、図7に示すグラフは表わしてい
る。

【0106】次に、話題転換尤度評価部115について
説明する。話題転換尤度評価部115は、3層のニュー
ラルネットで構成されている。入力層に与えられるデー
タと、各データの取り得る値は以下のとおりである。

【0107】第1は、キーワードの有無である。この
キーワードは、なしの場合0の値を取り、ありの場合
1の値を取る。このようなキーワードの有無は、前述
したように話題転換要因抽出部107によって話題転換
尤度評価部115に与えられる。

【0108】第2は、疑問文か否かである。疑問文であ
る場合は1を取り、それ以外は0を取る。このような疑
問文か否かは、話題転換要因抽出部107によって与え
られる。

【0109】第3は、相手特定語の有無である。対話の
相手を特定する語がない場合には0を取り、ある場合に
は1を取る。このような対話の相手を特定する語の有無
は、話題転換要因抽出部107によって与えられる。

【0110】第4は、同義語の数/有無である。同義語
(同語)が1発話または1文中に含まれない場合には0
を取り、含まれる場合には1またはその個数の値を取
る。このような同義語(同語)が1発話または1文中に
含まれている数は、同義語辞書125で得られた同義語
リストから同義語の数がカウントされることで得られ
る。

【0111】第5は、指示語の数/有無である。指示語
がない場合には0を取り、ある場合には1またはその個
数の値を取る。このような指示語の数は、単語辞書21
から得られた単語品詞リストから指示語が数えられるこ
とで得られる。

【0112】第6は、接続詞の数/有無である。接続詞
がない場合には0を取り、ある場合には1またはその個
数の値を取る。このような接続詞の数は、単語辞書21

から得られる単語品詞リストの接続詞が取出されること
で得られる。

【0113】第7は、初出語の数/有無である。初出語
がない場合には0を取り、ある場合には1またはその個
数の値を取る。このような初出語の有無や数は単語辞書
21から与えられる単語品詞リストが解析情報統計処理
部11で統計処理されることで得られる。

【0114】第8は、話題候補語の数/有無である。話
題候補語がない場合には0を取り、ある場合には1また
はその個数の値を取る。話題候補語の数や有無は、話題
候補語抽出部109で得られた話題候補語リストから話
題候補語が抽出されて得られる。

【0115】第9は、初出語かつ話題候補語の語の有無
/数である。初出語かつ話題候補語である語がない場合
には0を取り、ある場合には1またはその個数の値を取
る。このような初出語かつ話題候補語である語の有無や
数は、初出語の数/有無と話題候補語の数/有無より得
られる。

【0116】第10は、前文または前発話に含まれた語
の対比語の数/有無である。前文または前発話に含ま
れた語の対比語がない場合には0を取り、ある場合には1
またはその個数の値を取る。このような前文または前発
話に含まれた語の対比語の数や有無は、単語品詞リスト
から統計処理されて得られた情報と、現発話から出力さ
れた対比語リストとが比べられることで、対比語リスト
に含まれる語は前発話に現れているかどうか調べられ
て得られる。

【0117】第11は、各初出語の転換寄与指数であ
る。この話題転換寄与指数は0以上1以下の実数を取
り、初出語がない場合には0を取る。

【0118】第12は、各話題候補語の転換寄与指数で
ある。この話題転換寄与指数は0以上1以下の実数を取
り、候補語がない場合には0を取る。

【0119】第13は、話者の交代の有無である。話者
監視装置119で得られる話者IDと解析情報統計処理
部11で統計処理した統計情報から得られる。話者が交
代している場合には1を取り、交代していない場合には
0を取る。なお、この話者が交代したか否かに関して
は、処理が発話単位で行なわれれば通常は必ず1とな
る。

【0120】第14は、前話題転換からの経過時間であ
る。話題転換からの経過時間も0以上の実数を取る。こ
れは、話題転換尤度評価部115が転換を検出したとき
に、統計情報から前の話題転換検出發話の発話時刻が発
話推移時間監視装置121からの発話開始および終了時
刻から得られるので、現発話の発話開始時刻および終了
時刻とが演算されてその間の経過時間により得られる。

【0121】第15は、現発話者の前発話終了からの時
間である。この値は0以上の実数を取る。現発話の話者
IDと同じ話者IDの話者による前発話の開始時刻を統

計情報から得ることができ、現発話の開始時刻との差が解析情報統計処理部11で統計処理されることで経過時間が得られる。

【0122】第16は、現発話の発話時間である。この現発話の発話時間は、発話推移時間監視装置121から与えられる現発話の開始時刻および現発話の終了時刻の差から得られる。

【0123】第17は、前発話終了から現発話開始までの時間である。この時間は、発話推移時間監視装置121から与えられる現発話の開始時刻と、統計情報として記録されている前発話の終了時刻との差から得られる。

【0124】以上のような値が入力層に与えられると、出力層には話題転換の可能性が実数として出力される。最終的には、この出力される値に対して何らかの足切りが行なわれ、転換があったか否かが判定される。このように、話題転換の判断に使用される情報は、基本的に現在の発話から得られたものである。各種経過時間情報、語の出現頻度および対比語検出などで過去の発話の情報が使用されるが、従来の話題スタックによる話題同定法などとは異なる。すなわち、話題同定／転換の認識を遡って行なうためにそのような過去のデータが使用されるのではなく、あくまでも現在の発話で転換が発生したかどうかの判定のために使用される。これによって、本装置の実時間性は損なわれない。

【0125】話題転換度の評価は、必ずしもこの実施例のようにニューラルネットが使用される必要はない。たとえば、何らかの妥当な多項式などによる評価関数が定義されれば、それが使用されても構わない。ただし、判定に用いられる要素が上記のように非常に多く、しかも各要素の重み付けが困難であり、主観評価実験の結果から解析的に何らかの評価加数が獲得されることは、一般的に極めて困難であると思われる。そこで、ニューラルネットが用いられて主観評価実験から得られた結果が教師データとして与えられて学習が行なわれれば、比較的容易に評価関数に相当するものが得られる。そのために、本実施例ではニューラルネットが使用されている。

【0126】学習は、以下のように行なわれる。まず、ある対話文を5名の被検者に粗読してもらう。そして、話題転換があったと思った箇所にチェックをしてもらう。このようにして得られた結果から、2名以上の被検者が転換ありと判断した発話または文が話題転換のあった発話または文とされる。

【0127】次に、ニューラルネットの入力層に、各発話または文から得られる上記データが入力され、その発話または文で話題転換があった場合には1、なかった場合には0が出力層に与えられる。そして、バックプロパゲーションによって学習が行なわれる。このような学習が対話の初めから終わりまで順次繰返され、必要であれば同じ対話について複数回学習が行なわれる。これによ

うに学習される。または、 $\sin(x\pi/2m)$ のような関数（ここで、 x は話題転換を検出した被検者の数であり、 m は被検者総数である。）が用いられて教師データが生成される方法も考えられる。このような関数が用いられれば、総被検者数が少ない場合には教師データの転換尤度をより低く、逆にある程度以上の検出者がいる場合には全員が検出しなくても転換尤度をほぼ1にすることができ、より現実的な学習の可能性がある。

【0128】なお、実施例では被検者で2名以上が転換と認めた箇所を転換ありとして学習が行なわれている。しかし、他にも、たとえば5名のうち何名が転換ありと認識したかの割合を使って学習する方法も考えられる。たとえば、5名による主観評価実験で3名が転換ありとした箇所については、出力層に与える教師データを0.6とするような方法によってもよい。

【0129】このような方法によって、話題転換があったかどうか、時間遅れなく評価が可能となる。

【0130】次に、話題語同定部117について説明する。話題語同定部117は、話題転換尤度評価部115から話題の転換ありと通知されたときに、話題候補語抽出部109から与えられる現在の発話に含まれる話題候補語から話題と見なされる語を抽出し、使用者に提示する。話題候補語から話題と見なされる語が抽出される方法としては、以下のようなやり方が挙げられる。

【0131】図8は、そのような話題候補語から話題と見なされる語が抽出される種々の方法を説明するための図である。特に、図8には、図8(a)から図8(e)に示すように、以下のような5種類の方法を示している。

【0132】第1は、図8(a)に示すように、話題候補語すべてを話題候補語と見なして使用者に提示する方法である。

【0133】第2は、図8(b)に示すように、話題候補語で初出語があればそれを話題語として提示する。初出語でなければ、前回の出現から最も長い時間出現していなかった語を話題語として提示しておけばよい。または、初出語がなければ、前回の出現から長い時間出現しなかった順に所定の数の語を話題語として提示してもよい。さらに、初出語がなければ、出現頻度の最も低い語を話題語として提示してもよい。さらに、初出語がなければ、出現頻度の低いものから順に所定の数の語を話題語として提示してもよい。

【0134】第3の方法は、図8(c)に示すように、話題候補語から、過去に話題語として同定された語を削除した残りの語を話題語として提示する方法である。このような第3の方法の場合には、話題語同定結果である話題語が解析情報統計処理部111にフィードバックされることで、過去に話題語として同定された語に関する情報が解析情報統計処理部11に与えられる。

【0135】第4は、図8(d)に示すように、転換が

検出された発話内に複数回使用されている語、または同義語が出現する語が強調されていると見なして、これを話題語として提示する方法である。

【0136】第5は、図8(e)に示すように、話題転換要因抽出部107から、直前の発話内に含まれる語の対比語は本発話内に含まれることが通知されていれば、その対比語を話題語として提示する方法である。

【0137】以上のような種々の方法があるが、これらの方法が単独で用いられるだけでなく、いくつかが組合される方法でもよい。ただし、本装置では、話題の同定はあまり重要な位置を占めていない。すなわち、本装置の最も重要な目的は実時間に話題の転換を検出することであり、その検出に当たって予め話題が同定される必要がない。このことは、従来の技術と明らかに異なる。従来の技術では、まず話題の同定が行なわれ、その話題が同定された後に、ある部分の話題が変わっていれば転換があったと認識される。つまり、初めに話題があることが前提となる方法である。これに対し、本装置はその逆で、初めに話題の転換が前提とされる方法であり、転換が検出された後に必要であれば話題の同定が行なわれる。

【0138】ところで、話題の同定が行なわれることは現実には難しい。前述のような何らかの重み付けによって話題語が抽出されることは可能であるが、本当にその語が意味的に正しい話題を表すものかどうかの判断は、困難である。その正確な判断のために、「発話の意味理解」の機構が必要であり、特に自由展開対話型における発話内容の理解の可能な装置の構築は現状不可能と思われる。このため、従来の手法では、自由展開対話を対象としつつも、実際には分野辞書が準備される必要があり、この分野を外れた内容の発話は処理できない。つまり、話題の同定が行なわれてから、話題転換を検出するためには、発話の意味理解が一般的に不可欠であり、この結果装置は対応する分野ごとに何らかのドメイン知識が準備される必要がある。したがって、装置の適用可能な分野は自ずと限られる。

【0139】一方、本実施例の装置では、分野に依存する部分なくまず転換の認識が行なわれる。そして、転換の検出された発話に必ず次の話題が含まれる。極端に言えば、本装置では前述のような話題語同定が行なわれる必要もない。その場合、転換の検出された発話がそっくりそのまま使用者に提示されてもよい。つまり、発話内容の理解という困難な作業は、使用者に任せてもよく、このほうが、装置で中途半端な意味理解が行なわれるよりも確実に意味が伝わると思われるからである。

【0140】

【発明の効果】以上のように、この発明によれば、入力

された対話者の発話内容の形態素を解析し、予め定められた話題転換要因規則をその形態素の解析結果に適用して話題転換要因を抽出し、予め定められた話題規則をその形態素の解析結果に適用して話題候補語を抽出し、その後形態素の解析結果および話題候補語の抽出結果を統計処理するとともに、話題候補語の転換寄与指数を求め、話題転換要因、統計処理結果および話題候補語の転換寄与指数を評価して、発話内容に話題転換が存在するかどうかを極力実時間で検出できる。

【0141】さらに、目的指向型または自由展開型のような対話の型や話題の分野の問うことなく、話題転換が検出された場合の話題同定をも行なえる。

【図面の簡単な説明】

【図1】この発明の一実施例による対話処理装置の概略ブロック図である。

【図2】この発明の他の実施例による対話処理装置を示した概略ブロック図である。

【図3】図2の解析情報統計処理部の動作を説明するための第1の図である。

【図4】図2の解析情報統計処理部の動作を説明するための第2の図である。

【図5】図2の解析情報統計処理部の動作を説明するための第3の図である。

【図6】図2の単語の話題転換寄与指数演算部の動作で用いられる関数を示したグラフである。

【図7】図6に示された関数に対して、実験を行なった結果を示したグラフである。

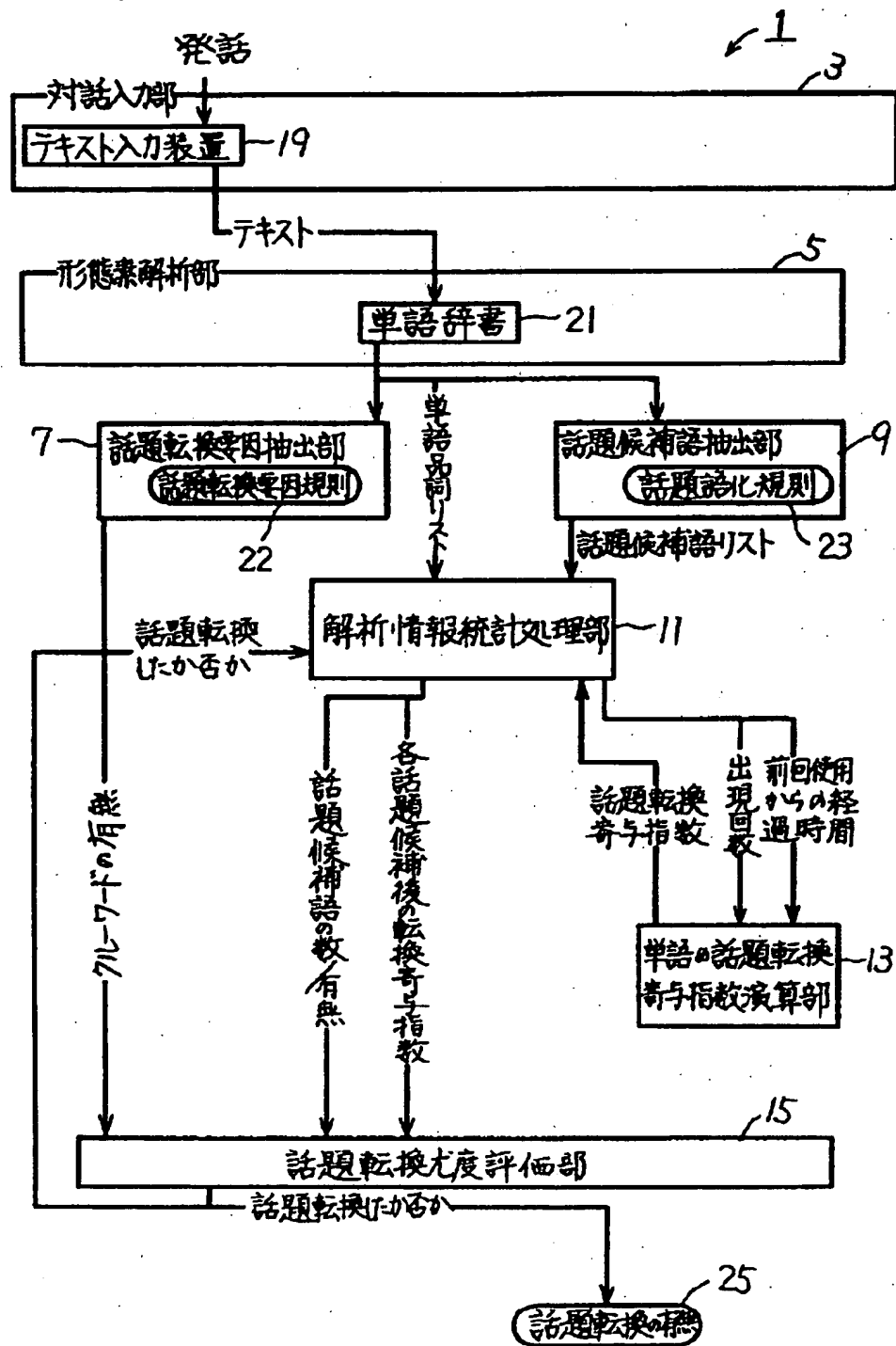
【図8】図2の話題語同定部の動作を説明するための図である。

【図9】従来の話題理解の方法を説明するための図である。

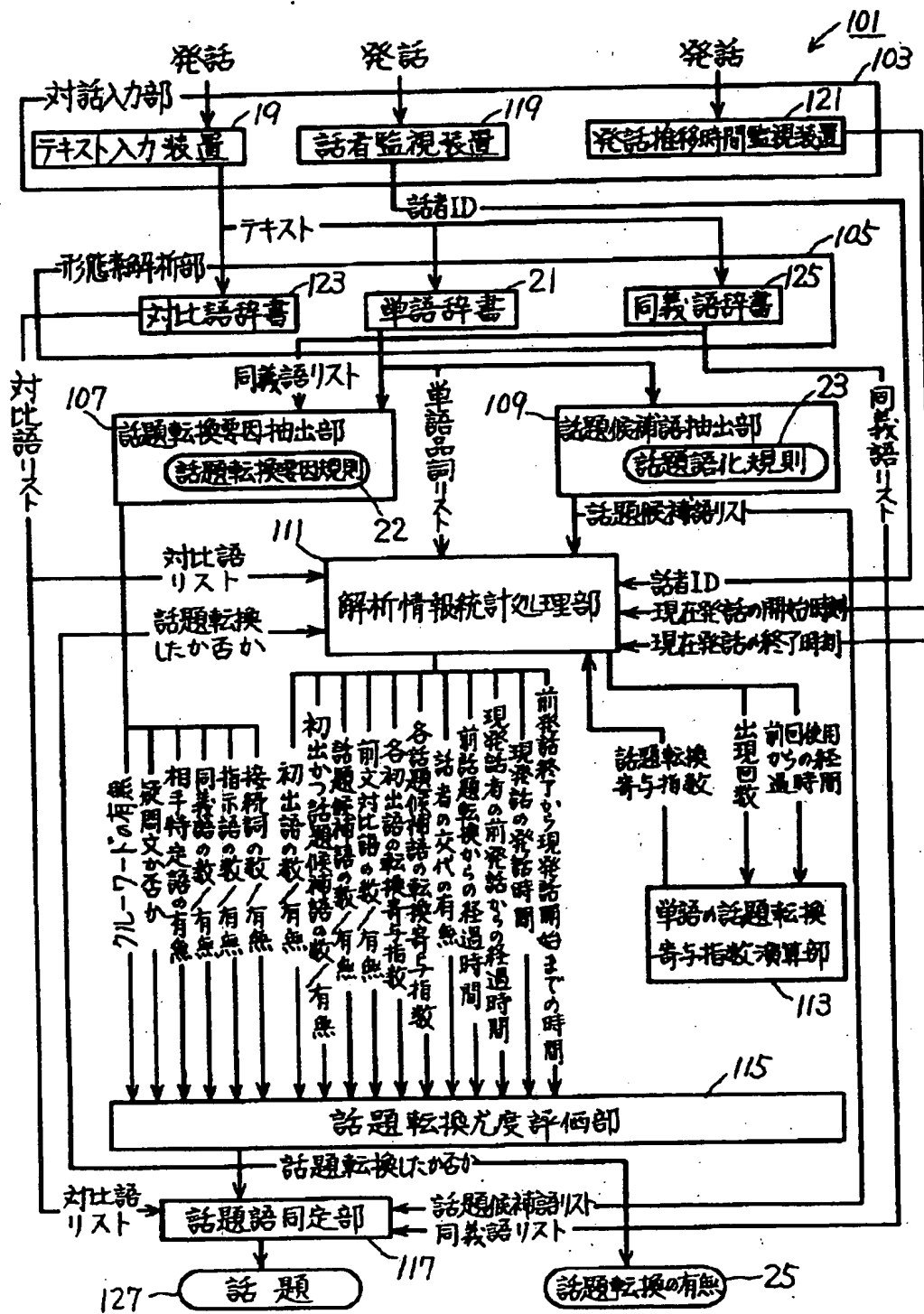
【符号の説明】

- 1, 101 対話処理装置
- 3, 103 対話入力部
- 5, 105 形態素解析部
- 7, 107 話題転換要因抽出部
- 9, 109 話題候補語抽出部
- 11, 111 解析情報統計処理部
- 13, 113 単語の話題転換寄与指数演算部
- 15, 115 話題転換尤度評価部
- 117 話題語同定部
- 19 テキスト入力装置
- 119 話者監視装置
- 121 発話推移時間監視装置
- 123 対比語辞書
- 21 単語辞書
- 125 同義語辞書

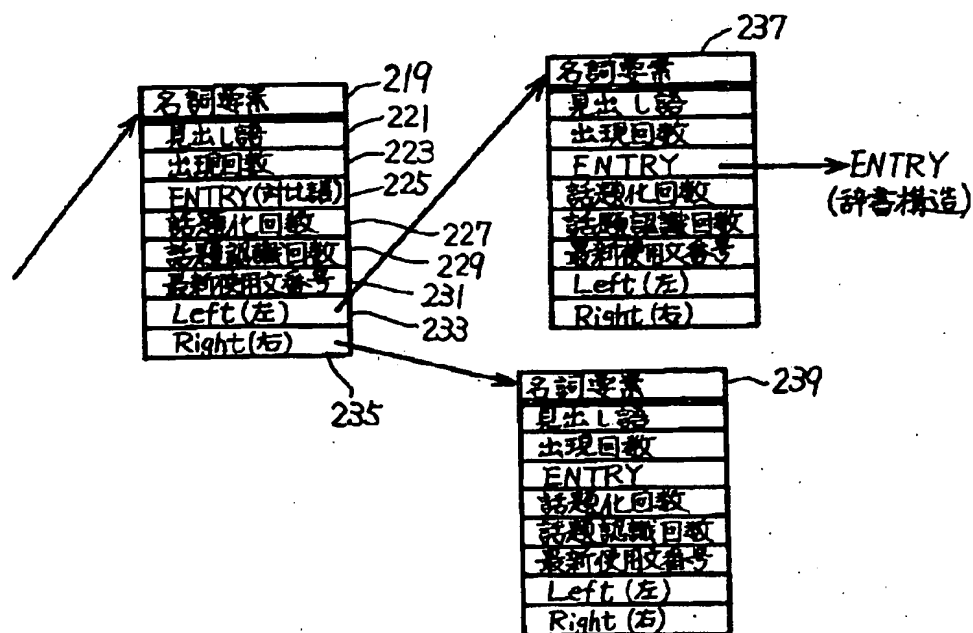
【図1】



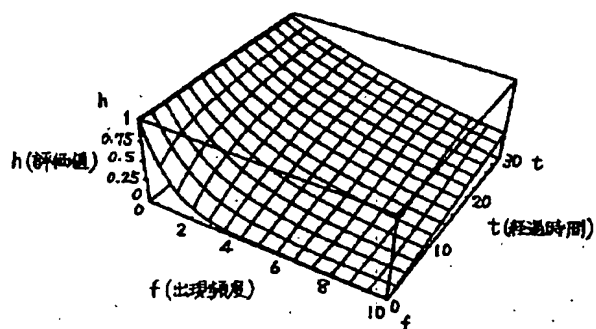
【図2】



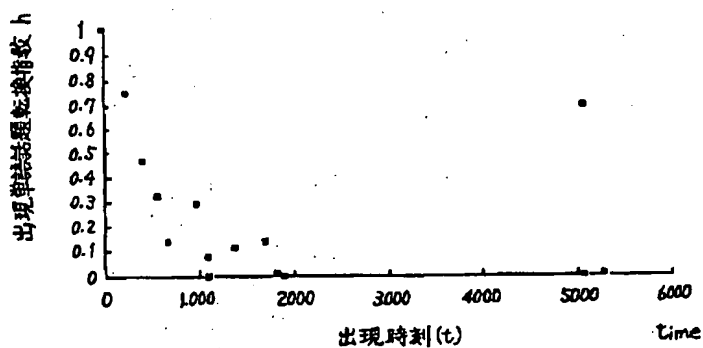
【図4】



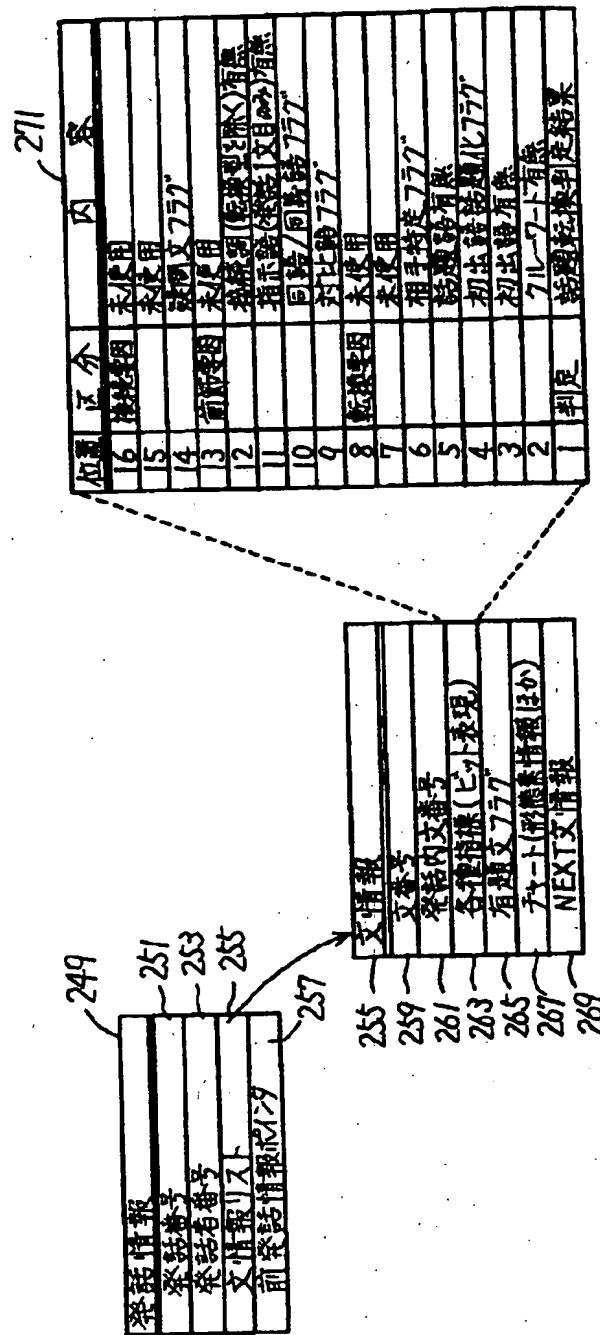
【図6】



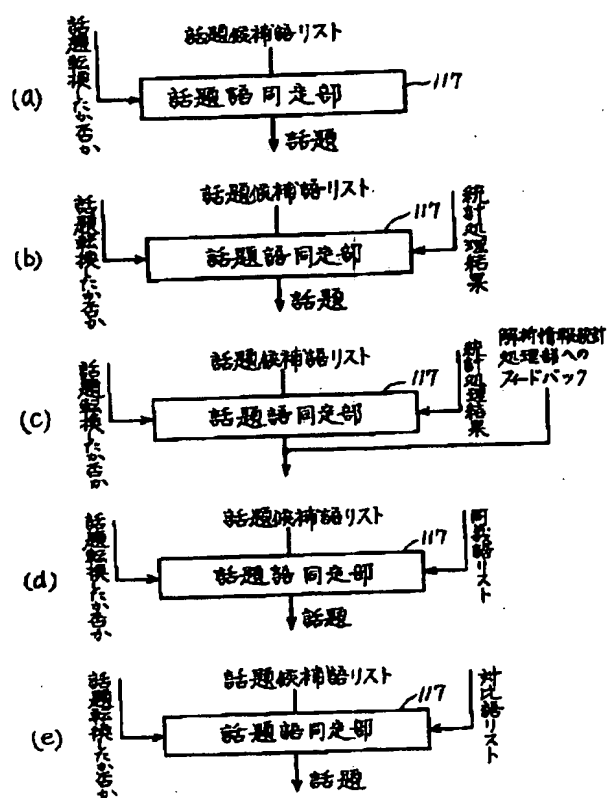
【図7】



【図5】



【图 8】



【図9】

